Journal of Research in Science and Engineering (JRSE)
ISSN: 1656-1996 Volume-7, Issue-1, January 2025
DOI: 10.53469/jrse.2025.07(01).13

Machine Learning Methods in Detecting Heart Related Diseases

Parul Khanna

School of Computer Science, SYMCA, Dr. Vishwanath Karad Maharashtra, Institute of Technology, World Peace University.
Pune, India
khanna28@gmail.com

Abstract: ***Cardiovascular illnesses, often known as heart - related disorders or CVDs, have been the leading cause of mortality worldwide in recent decades and are now the most serious illness, not just in India but around the world. Therefore, a system that is dependable, accurate, and workable is required to identify these illnesses in time for appropriate treatment. Large - scale and sophisticated data processing has been automated by using machine learning techniques and algorithms to a variety of medical datasets. Recently, a number of researchers have started employing various machine learning approaches to assist the medical community and experts in the detection of heart - related illnesses. An overview of several models built using these methods and algorithms is presented in this work.***

Keywords: Cardiovascular Diseases; Support Vector Machines; K - Nearest Neighbour; Naïve Bayes; Decision Tree

1. Introduction

The heart is an important organ in the human body. It pumps blood to every part of our anatomy. If it does not work properly, the brain and many other organs stop working and the person dies within minutes. Changes in lifestyle, work stress and poor eating habits increase the prevalence of several heart diseases. Heart disease has become one of the leading causes of death worldwide. According to the World Health Organization, heart disease kills 17.7 million people every year, which is 31% of all global deaths. Heart disease has also become the leading cause of death in India [1]. According to the Global Burden of Disease Report 2016 released on September 15, 2017, 1.7 million Indians died of heart disease in 2016. Heart disease increases healthcare costs and also reduces a person's productivity. According to the World Health Organization (WHO), India lost as much as $237 billion due to heart - related or cardiovascular diseases between 2005 and 2015 [2]. Therefore, feasible and accurate prediction of heart disease is very important. Medical organizations around the world collect information on various health - related issues. This information can be used by various machine learning techniques to provide useful insights. However, the data collected is very massive and often this data can be very noisy. These datasets, which are too large for the human mind to process, can be easily explored using various machine - learning techniques. Thus, these algorithms have recently become very useful in accurately predicting the presence or absence of heart disease.

2. Dimensionality Reduction

Dimensionality reduction involves choosing a mathematical representation in such a way that most, but not all, of the variance in the given data can be accounted for, thus containing only the most relevant information. The information of a task or problem can consist of several attributes or dimensions, but not all these attributes can affect the output equally. A large number of attributes or features can affect the computational complexity and even lead to over - tuning, leading to poor results. So, size reduction is a very important step in building any model. Dimensionality reduction is generally achieved by two methods – feature removal and feature selection.

a) Feature Extraction

Here, the new set is derived from the original set. Feature extraction involves feature transformation. This change often cannot be reversed, as some or perhaps much of the useful information is lost in the process. [3] and [4] use principal component analysis (PCA) for feature extraction. Principal component analysis is a commonly used linear transformation algorithm. In feature space, it finds cues that maximize variance and finds cues that are perpendicular to each other. It is a global algorithm that provides the best reconstruction.

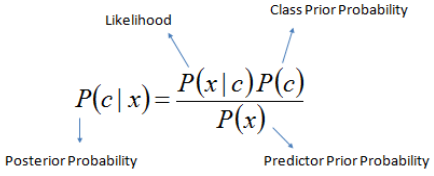
b) Feature Selection

Here a subset of the original set is selected. In [5], essential features are selected by CFS (correlation - based feature selection) subset evaluation combined with a best - first search method to reduce dimensionality. [6] The chi - square statistical test selects the most important features.

3. Algorithms and Techniques Used

a) Naïve Bayes

Naive Bayes is a simple but powerful classification technique, based on Bayes' theorem. This assumes independence of\predictions, i. e. attributes or properties do not correlate with each other or be related to each other in any way. Even if there is a dependency, all these features or attributes affect the probability independently and hence it is called naive.

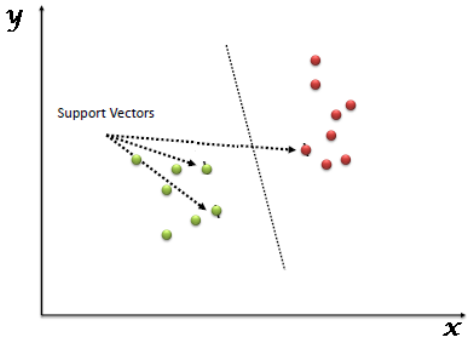


$$P(c \mid \mathrm{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

In [7], Naive Bayes achieved 84.1584% accuracy with the top 10 features selected by SVMRFE (Recursive Feature Elimination) [8] Naive Bayes achieved 83.49% accuracy using all 13 features of the Cleveland dataset [25].

b) Support Vector Machine

A Support Vector Machine is a very popular supervised machine learning technique (with a predefined target variable), which can be used as a classifier and predictor. For classification, it finds a hyperplane in the feature space that differentiates the classes. The SVM model represents the training data points as points in the object space, mapped so that points belonging to different classes are separated with the widest possible margin. The test data points are then mapped to the same space and classified according to which side of the margin they belong.



Shan Xu et al. used SVM to achieve 98.9% accuracy on the People's Hospital dataset [5]. In [9], SVM performs the best, 85.7655% of cases are correctly classified, and [10] uses the SVM boosting technique gives 84.81% accuracy. Houda Mezrigui et al. used SVM to achieve an f - measure value of 93.5617 [11]. In [12], SVM classifies pixel variations with 92.1% accuracy, which helps in detecting the damaged area inaccurately.

c) K – Nearest Neighbour

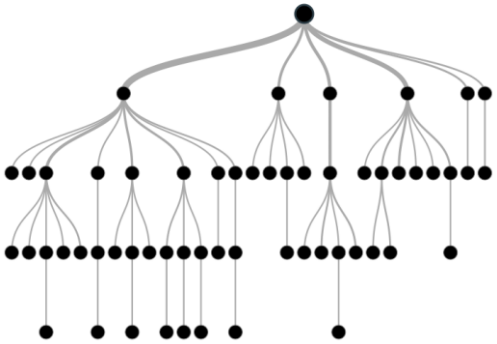
In 1951, Hodges et al. introduced a non - parametric technique for pattern classification commonly known as the K - nearest - neighbour rule [13]. The K - Nearest Neighbour technique is one of the most basic but very effective classification methods. It makes assumptions about the data and is typically used for classification tasks when there is little or no prior knowledge about the distribution of the data. This algorithm involves finding the closest data points in the training set to a data point for which the target value is not available and putting the mean of the data points found in it. [10] KNN provides accuracy. of 83.16% when k is equal to 9 when the 10 - cross - check technique is used. [14] KNN and Ant Colony Optimization outperform other techniques with an accuracy of 70.26% and an error rate of 0.526. Ridhi Saini et al. achieved an efficiency of 87.5% [15], which is very good.

d) Decision Tree

A decision tree is a supervised learning algorithm. This technique is mostly used to solve classification problems. It works effortlessly with continuous and categorical functions. This algorithm divides the population into two or more similar sets based on the most important predictors. The decision tree algorithm first calculates the entropy of each attribute. The dataset is then partitioned using the variables or predictors with the highest informative gain or lowest entropy. These two operations are performed recursively with the other properties.

$$Entropy(S) = \sum_{i=1}^{c} -p_i \log_2 p_i$$

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$



In [10] A choice tree has the most exceedingly bad execution with an accuracy of 77.55% but when a choice tree is utilized with a boosting technique performs superior with a precision of 82.17%. In [9] choice performs exceptionally ineffectively with an accurately classified instance percentage of 42.8954% though [16] moreover employs the same dataset but utilises the J48 calculation for actualizing the Decision Trees and the precision in this way gotten is 67.7% which is less but still a change on the previous. Renu Chauhan et al. have obtained an exactness of 71.43% [17]. M. A. Jabbar et al. have used alternating choice trees with guideline component examination to obtain a precision of 92.2% [18]. Kamran Farooq et al. have achieved the best comes by employing a choice tree - based classifier combined with a forward determination which accomplishes a weighted precision of 78.4604% [19].

4. Conclusion

Based on the above review, it can be concluded that machine learning algorithms have a safe potential to predict cardiovascular or heart diseases. All of the above algorithms performed very well in some cases but poorly in others. Alternative decision trees performed very well when used with PCA, but decision trees performed very poorly in some other cases, which may be due to overfitting. The Random Tree Forest and Ensemble models performed very well because they solved the problem of all matching using multiple algorithms (in the case of a random forest, multiple decision trees). Models based on the naive Bayes classifiers were computationally very fast and worked well. SVM performed very well in most cases. Systems based on machine learning algorithms and techniques have been very accurate in predicting heart disease, but much research remains to be done on how to handle high - dimensional data and overfitting. Much research can also be done on the right set of algorithms to use for certain types of data.

References

[1] Ramadoss and Shah B et al. “A. Responding to the threat of chronic diseases in India”. Lancet.2005; 366: 1744–1749. doi: 10.1016/S0140 - 6736 (05) 67343 - 6.

[2] Global Atlas on Cardiovascular Disease Prevention and Control. Geneva, Switzerland: World Health Organization, 2011

[3] Dhomse Kanchan B and Mahale Kishor M. et al. “Study of Machine Learning Algorithms for Special Disease Prediction using Principal of Component Analysis”, 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication.

[4] R. Kavitha and E. Kannan et al. “An Efficient Framework for Heart Disease Classification using Feature Extraction and Feature Selection Technique in Data Mining “, 2016

[5] Shan Xu, Tiangang Zhu, Zhen Zang, Daoxian Wang, Junfeng Hu and Xiaohui Duan et al. “Cardiovascular Risk Prediction Method Based on CFS Subset Evaluation and Random Forest Classification Framework”, 2017 IEEE 2nd International Conference on Big Data Analysis.

[6] Manpreet Singh, Levi Monteiro Martins, Patrick Joanis and Vijay K. Mago et al. “Building a Cardiovascular Disease Predictive Model using Structural Equation Model & Fuzzy Cognitive Map”, 978 - 1 - 5090 - 0626 - 7/16/$31.00 c 2016 IEEE.

[7] Kanika Pahwa and Ravinder Kumar et al. “Prediction of Heart Disease Using Hybrid Technique For Selecting Features”, 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON).

[8] Seyedamin Pouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia, Juan Gutierrez et al. “ A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease”, 22nd IEEE Symposium on Computers and Communication (ISCC 2017): Workshops - ICTS4eHealth 2017

[9] Hanen Bouali and Jalel Akaichi et al. “Comparative study of Different classification techniques, heart Diseases use Case. ”, 2014 13th International Conference on Machine Learning and Applications

[10] Seyedamin Pouriyeh, Sara Vahid, Giovanna Sannino, Giuseppe De Pietro, Hamid Arabnia, Juan Gutierrez et al. “ A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease”, 22nd IEEE Symposium on Computers and Communication (ISCC 2017): Workshops - ICTS4eHealth 2017

[11] Houda Mezrigui, Foued Theljani and Kaouther Laabidi et al. “Decision Support System for Medical Diagnosis Using a Kernel - Based Approach”, ICCAD’17, Hammamet - Tunisia, January 19 - 21, 2017.

[12] Dr. (Mrs). D. Pugazhenthi, Quaid - E - Millath and Meenakshi et al. “Detection Of Ischemic Heart Diseases From Medical Images “ 2016 International Conference on Micro - Electronics and Telecommunication Engineering.

[13] J. Hodges et al. “Discriminatory analysis, nonparametric discrimination: Consistency properties, ” 1981.

[14] S. Rajathi and Dr. G. Radhamani et al. “Prediction and Analysis of Rheumatic Heart Disease using kNN Classification with ACO “, 2016.

[15] Puneet Bansal and Ridhi Saini et al. “Classification of heart diseases from ECG signals using wavelet transform and kNN classifier”, International Conference on Computing, Communication and Automation (ICCCA2015).

[16] Simge EKIZ and Pakize Erdogmus et al. “Comparitive Study of heart Disease Classification”, 978 - 1 - 5386 - 0440 - 3/17/$31.00 ©2017 IEEE.

[17] Renu Chauhan, Pinki Bajaj, Kavita Choudhary and Yogita Gigras et al. “Framework to Predict Health Diseases Using Attribute Selection Mechanism”, 2015 2nd International Conference on Computing for Sustainable Global Development (INDIA Com).

[18] M. A. JABBAR, B. L Deekshatulu and Priti Chndra et al. “Alternating decision trees for early diagnosis of heart disease”, Proceedings of International Conference on Circuits, Communication, Control and Computing (I4C 2014).

[19] Amir Hussain, Peipei Yang, Mufti Mahmud and Jan Karasek et al. “A Novel Cardiovascular Decision Support Framework for effective clinical Risk Assessment. ”, 978 - 1 - 4799 - 4527 - 6/14/$31.00 ©2014 IEEE.

[20] Quazi Abidur Rahman, Larisa G. Tereshchenko, Matthew Kongkatong, Theodore Abraham, M. Roselle Abraham, and Hagit Shatkay et al. “Utilizing ECG - based Heartbeat Classification for Hypertrophic Cardiomyopathy Identification”, DOI 10.1109/TNB.2015.2426213, IEEE Transactions on Nano Bioscience TNB - 00035 - 2015.